1025 (3:05) 6 "Always Late With Your Kisses", Lefty Frizzell, **The Best Of Lefty Frizzell** (Rhino R2 71005)

1026 (3:29) 2 "Always On My Mind", Willy Nelson, **Always On My Mind** (Columbia CK 37951)

1146 (2:58) 2 "She's Got You", Patsy Cline, **The Patsy Cline Story** (MCA CAD-4038)

CONTEMPORARY JAZZ- 6 29:20

1101 (4:58) 2 "Forever In Love", Kenny G., **Breathless** (Arista 18646)

1123 (5:57) 2 "Facing West", Pat Metheney, **Secret Story** (Geffen GEFD 24468)

1124 (6:07) 6 "Children's World", Maceo Parker, **Life On Planet Groove** (Verve 517 197)

1095 (5:40) 4 "Dewey", Yellowjackets, **Like A River** (GRP 9689)

1100 (4:03) 8 "Embrace", Richard Elliot, **Soul Embrace** (Manhattan 98946)

CLASSIC JAZZ- 7 29:07

1083 (11:31) 4 "All Blues", Miles Davis, **Kind Of Blue** (Columbia CK 40579)

1142 (8:03) 2 "Perdido" Dave Brubeck Quartet, **Jazz At Oberlin** (Fantasy F 3245)

1001 (4:27) 9 "Naima", John Coltrane, **Giant Steps** (Atlantic 1311-2)

1002 (4:59) 1 "Lonely Woman", Ornette Coleman, **The Shape Of Jazz To Come** (Atlantic 1317-2)

BLUES- 8 26:53

1027 (4:17) 2 "Boom Boom", John Lee Hooker, **Boom Boom** (Charisma V2-86553)

1028 (3:30) 3 "The Thrill Is Gone", B.B. King, **Live At The Apollo** (GRP GRD-9637)

1029 (3:31) 9 "Blues #572", Johnnie Johnson, **Johnnie B. Bad** (Elektra Nonesuch 961149-2)

1030 (4:40) 2 "I'd Rather Go Blind", Koko Taylor, **Live From Chicago** (Alligator Al 4754)

1031 (3:11) 21 "Cold, Cold Feeling", T-Bone Walker, **The Complete Imperial Recordings** (EMI CDP-7-96737-2)

1032 (4:47) 5 "There Is Something On Your Mind", Buddy Guy, **Damn Right, I've Got The Blues** (Silvertone 1462-2-J)

1179 (3:24) 1 "When I'm Drinkin", Champion Jack Dupree, **Back Home in NewOrleans**, (Bullseye Blues Company BB 9502)

BIG BAND/SWING- 9 29:27

1060 (3:25) 11 "Ain't Misbehavin," Tony Bennett/Count Basie, **Tony Bennett and Count Basie** (Laser 15 722)

1058 (3:23) 14 "Rhapsody in Blue," Glen Miller, **The Kings of Swing** (Laser 15 714)

1059 (3:11) 6 "Begin The Beguine," Artie Shaw, **Artie Shaw** (Laser 15 713)

1059 (3:29) 8 "Stardust," Artie Shaw, **Artie Shaw** (Laser 15 713)

1057 (1:37) 3 "I'll Get By," Billie Holiday, **The First Esquire Concert** (Laser 15 723)

1057 (8:38) 9 "I Got Rythm," Tatum/Armstrong, **The First Esquire Concert** (Laser 15 723)

1060 (2:15) 5 "Anything Goes," Tony Bennett/Count Basie, **Tony Bennett and Count Basie** (Laser 15 722)

TOP OF THE CHARTS- 10 27:44

1110 (4:32) 1 "Little Bird", Annie Lennox, **Diva** (Arista 07822-18704-2)

1114 (4:06) 1	"A Whole New World (Aladdin's Theme), Peabo Bryson & Regina Belle, **Aladdin Original Motion Picture Soundtrack** (Disney 60846-2)
1111 (3:32) 1	"Walk On The Ocean", Toad The Wet Sprocket, **Fear** (Columbia 47309)
1112 (5:45) 1	"Keep The Faith", Bon Jovi, **Keep The Faith** (Jamco/Mercury 314 514 045-2)
1090 (4:28) 1	"Tears In Heaven", Eric Clapton, **Unplugged** (Duck/Reprise 45024)
1113 (4:30) 1	"I Will Always Love You", Whitney Houston, **Bodyguard Original Soundtrack** (Arista 07822-18699-2)

<u>CLASSIC ROCK- 11</u> 26:48

1003 (4:35) 4	"Gimme Shelter", The Rolling Stones, **Let It Bleed**, (ABKCO 80042)
1004 (3:13) 4	"Don't Stop", Fleetwood Mac, **Rumours**, (Warner Bros. 3010-2)
1005 (4:58) 1	"Baba O'Riley", The Who, **Who's Next** (MCA 37217)
1006 (2:20) 3	"Fortunate Son", Credence Clearwater Revival, **Chronicle Vol. 1**(Fantasy CCR-2)
1007 (5:02) 2	"Bell Bottom Blues", Derek & The Dominoes, **Layla** (Polydor 847 090-2)
1144 (3:23) 1	"Refugee", Tom Petty, **Damn The Torpedoes** (MCA 31161)
1138 (3:37) 3	"Changes", David Bowie, **Changes** (Rykodisc RCD 20171)

<u>FIFTIES- 12</u> 26:09

1008 (2:13) 9	"I'm Walkin'", Fats Domino, **My Blue Heaven** (EMI E2792808-2)
1079 (2:29) 10	"Peggy Sue", Buddy Holly, **From The Original Master Tapes** (MCA 5540)

1072 (2:28) 9	"Jailhouse Rock", Elvis Presley, **The Number One Hits** (RCA 6382-2 R)
1077 (2:39) 9	"Breathless", Jerry Lee Lewis, **Original Sun Greatest Hits** (Rhino R2 70255)
1078 (2:17) 6	"Runaway", Del Shannon, **Little Town Flirt** (Rhino R2 70983)
1071 (2:06) 10	"Wake Up Little Susie", The Everly Brothers, **The Everly Brothers** (Rhino R2 70211)
1076 (1:56) 14	"Summertime Blues", Eddie Cochrane, **Legends Of Rock & Roll** (EMI E2792809)
1086 (2:08) 15	"Good Golly, Miss Molly", Little Richard, **Stand By Me (Original Soundtrack)** (Atlantic 7 81677-2)
1079 (2:01) 5	"Maybe Baby", Buddy Holly, **From The Original Master Tapes** (MCA 5540)
1072 (2:09) 1	"Heartbreak Hotel", Elvis Presley, **Number One Hits** (RCA 6382-2 R)
1071 (2:17) 1	"This little Girl of Mine", The Everly Brothers, **The Everly Brothers** (Rhino R2 70211)
SIXTIES- 13	29:12
1074 (2:42) 11	"Fa-Fa-Fa-Fa-Fa", Otis Redding, **Dictionary Of Soul** (Rhino 7 91707-2)
1009 (2:48) 8	"Crying", Roy Orbison, **For The Lonely: 18 Greatest Hits** (Rhino A 20574)
1075 (2:33) 6	"Hold On, I'm Coming", Sam & Dave, **Hold On, I'm Coming** (Atlantic 7 82055-2)
1010 (2:27) 1	"Respect", Aretha Franklin, **I've Never Loved A ManThe Way I Love You**, (Atlantic 8139 2)
1081 (2:11) 26	"Soothe Me", Sam Cooke, **The Man And His Music** (RCA PCD1-7127)

1073 (2:41) 7	"I Can't Help Myself", The Four Tops, **Greatest Hits** (Motown 3746352082)
1094 (2:46) 1	"I Got You", James Brown, **20 All-Time Greatest Hits** (Polydor 314 511 326-2)
1105 (2:34) 2	"Ain't Too Proud To Beg", The Temptations, **Great Songs That Inspired The Motown 25th Anniversary** (Motown MCD09033MD)
1133 (2:09) 13	"Yesterday", The Beatles, **HELP!** (EMI CDP 7 46439 2)
1134 (2:40) 1	"Be My Baby", The Ronettes, **The Best Of The Ronettes** (ABKCO 72122)
1143 (2:25) 1	"Wouldn't It Be Nice", The Beach Boys, **Pet Sounds** (Capitol 48421)

<u>FOLK ROCK- 14</u> 25:57

1011 (3:04)4	"Carey", Joni Mitchell, **Blue** (Reprise 2038-2)
1117 (2:55) 12	"Blowin In The Wind", Peter, Paul, & Mary, **In The Wind** (Warner Bros. 9 26224 2)
1012 (3:21) 1	"Picture Show", John Prine, **The Missing Years** (Oh Boy! 009)
1080 (4:04) 1	"Play Me Backwards", Joan Baez, **Play Me Backwards** (Virgin 86458-2)
1013 (3:08) 5	"Sealed With A Kiss", Lavin, Fingerett, Gold, McDonough, **"Buy Me Bring Me Take Me: Don't Mess My Hair..." Life According To Four Bitchin' Babes, Vol. 2** (Rounder/Philo 1150)
1014 (4:31) 5	"Sittin' On Top Of The World", Bob Dylan, **Good As I Been To You** (Columbia CK 53200)
1136 (4:00) 2	"Carolina In My Mind", James Taylor, **Greatest Hits** (Warner Bros. 31132)

LATIN BALLADS- 15 29:38

1150 (3:44) 6 "Atame A Tu Vida", Daniela Romo, De Mil Colores (Capitol

1147(3:58) 4 "Cree En Nuestro Amor", Jon Secada, **Otrodia Mas Sin Verte** (SBK/Capitol 80646)

1097 (4:00) 1 "Castillo Azul", Ricardo Montaner, **Los Hijos Del Sol** (Rodven 2995)

1148 (4:27) 11 "Exxtasis", Chayanne, **Provocame** (Sony CDZ 80831)

1106 (4:00) 1 "40 Y 20", Jose Jose, **40 Y 20** (Ariola

1104 (4:50) 2 "Cronica De Un Viejo Amor", Braulio, **Entre El Amor Y El Deseo** (Sony Latin

1153 (3:33) 1 "Mio", Paulina Rubio, **Paulina Rubio** (Capitol 42750 2)

LATIN RHYTHMS- 16 26:02

1152 (4:10) 2 "El Costo De La Vida", Juan Luis Guerra, **Areito** (Karen 146)

1149 (3:37) 2 "Menealo", Fransheska, **Menealo** (Ariola 3207 2)

1151 (3:50) 1 "Corazo Herido", Azucar Moreno, **Con La Miel En Los Labios** (Capitol/EMI 42721)

1154 (4:46) 2 "Salsa Y Control", Orquesta De La Luz, **Somos Diferentes** (Sony Discos 80851)

1152(4:08) 5 "Frio, Frio," Juan Luis Guerra, **Areito** (Karen 146)

1154 (4:44) 6 "Salsa Con Sabor", Orquesta De La Luz, **Somos Diferentes** (Sony Discos 80851)

REGGAE- 17 23:15

1036 (4:05) 6 "Murder, She Wrote", Chaka Demus & Pliers, **Bam Bam It's Murder** (Mango 162-539 923-2)

1033 (4:01) 4 "Muscle Grip", Shabba Ranks, **X-Tra Naked** (Epic 52464)

1037 (3:27) 8 "Satta Massagana", The Abyssinians, Satta Massagana (Heartbeat HB 120)

1038 (3:30) 9 "Two Sevens Clash", Culture, **Two Sevens Clash** (Shanachie SH 44001

1139 (4:27) 2 "Lee And Molly", Ziggy Marley And The Melody Makers, **Conscious Party** (Virgin 7 90878-2)

1140 (3:00) 4 "Three Little Birds," Bob Marley and the Wailers, **Legend** (Tuff Gong 422 846 210 2)

<u>RAP/HIP-HOP- 18</u> 25:52

1035 (4:19) 6 "Rebirth Of Slick (Cool Like That)", Digable Planets, **Reachin' (A New Refutation Of Time And Space)** (Pendulum 64674)

1039 (3:55) 4 "Wicked", Ice Cube, **The Predator** (Priority P2-57185)

1040 (3:52) 9 "I Got A Man", Positive K, **Skills That Pay The Bills** (Island 314 514 057 2)

1041 (4:06) 2 "Who's The Man?", Heavy D. & The Boyz, **Blue Funk** (Uptown 10734)

1042 (4:43) 6 "New Jack Hustler", Ice-T, **O.G. Original Gangster** (Sire/Warner Bros. 9 26492-2)

1129 (4:06) 5 "Mr. Wendel", Arrested Development, **Three Years, Five Months, And Two Days In The Life Of...** (Chrysalis F2-21929)

<u>DANCE- 19</u> 29:11

1085 (5:17) 4 "Get Away", Bobby Brown, **Bobby** (MCA MCAD 10417)

1062 (5:24) 2 "Reminisce", Mary J. Blige, **What's The 411?,** (Uptown UPTD 10681)

1033 (4:04) 4 "Muscle Grip", Shabba Ranks, **X-Tra Naked** (Epic 52464)

1064 (5:33) 4	"Deeper And Deeper", Madonna, **Erotica** (Maverick/Sire 45031-2)
1034 (4:02) 5	"Sunshine And Love", Happy Mondays, **...Yes Please!** (Elektra 61391)
1035 (4:22) 6	"Rebirth of Slick (Cool Like Dat)", Digable Planets, **Reachin' (A New Refutation Of Time And Space)** (Pendulum 64674)

<u>SONGS OF LOVE- 20</u> 26:46

1118 (2:38) 1	"Strangers In The Night", Frank Sinatra, **Greatest Hits** (Reprise 2274 2)
1119 (5:27) 1	"I've Got You Under My Skin", Dinah Washington, **Cole Porter Songbook Night And Day** (Verve 847202-2)
1119 (3:09) 16	"Night And Day", Ella Fitzgerald, **Cole Porter Songbook Night And Day** (Verve 847202-2)
1119 (3:48) 14	"In The Still Of The Night", Billy Eckstine, **Cole Porter Songbook Night And Day** (Verve 847202-2)
1119 (2:59) 15	"Easy To Love", Billie Holiday, **Cole Porter Songbook Night And Day** (Verve 847202-2)
1108 (3:54) 7	"I Can't Give You Anything But Love," Judy Garland, **Greatest Hits** (Curb D2 77370)
1126 (2:20) 1	"The Sea of Love", Phil Phillips, **The Sea Of Love Original Motion Picture Soundtrack** (Mercury 842 170-2)
1118 (1:54) 10	"When Somebody Loves You", Frank Sinatra, **Greatest Hits** (Reprise 2274 2)

<u>SINGERS + STRINGS- 21</u> 27:48

1115 (3:03) 9	"Crazy He Calls Me", Billie Holiday, **From The Original Decca Masters** (MCA MCAD 5766)
1108 (3:30) 6	"Over The Rainbow", Judy Garland, **Greatest Hits** (Curb D2 77370)

1122 (3:13) 12 "Blue Moon", Ella Fitzgerald, **Rodgers And Hart Songbook Vol 2** (Verve 821 579 2)

1109 (2:33) 6 "Love Me Love Me", Dean Martin, **Collector Series** (Capitol 91633-2)

1119 (6:14) 13 "I Get A Kick Out Of You", Dinah Washington, **Cole Porter Songbook Night And Day** (Verve 847202-2)

1118 (4:25) 3 "It Was A Very Good Year", Frank Sinatra, **Greatest Hits** (Reprise 2274 2)

1119 (3:46) 14 "In The still of The Night," Billy Eckstein, **Cole Porter Songbook Night And Day** (Verve 847202-2)

INSTRUMENTALS 22 28:56

1125 (2:54) 9 "Georgy Girl", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (2:44) 1 "All The Way" 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (3:46) 2 "Evergreen", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (7:08) 3 "Mac Arthur Park", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (4:10) 4 "Down By The Riverside", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (2:22) 5 "Put Your Head On My Shoulder", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (2:31) 6 "Tie A Yellow Ribbon", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

1125 (3:18) 7 "The Way We Were", 101 Strings, **More Hits Of The 50's-60's-70's** (Alshire ALCD 36)

HEAVY METAL- 23 27:44

1070 (4:15) 2 "In Bloom", Nirvana, **Nevermind** (DGC DGCD-24425)

1068 (5:42) 2 "Sad But True", Metallica, **untitled** (Elektra 961113-2)

1069 (5:04) 5 "Sweating Bullets", Megadeth, **Countdown To Extinction** (Capitol C2-98531)

1067 (3:58) 12 "Highway To Hell", AC/DC, **AC/DC Live** (Atlantic 92215-2)

1046 (4:05) 1 "Bombtrack", Rage Against The Machine, **Rage Against The Machine** (Epic 52959)

1132 (4:12) 3 "Mama, I'm Coming Home", Ozzy Osbourne, **No More Tears** (Epic Associated ZK 46795)

ALBUM ROCK- 24 26:44

1066 (3:40) 1 "Off The Ground", Paul McCartney, **Off The Ground** (Capitol 80362)

1065 (4:18) 7 "Two Princes", Spin Doctors, **Pocket Full Of Kryptonite** (Epic ZK 47461)

1043 (5:14) 10 "Man On The Moon", R.E.M., **Automatic For The People** (Warner Bros. 9 45055-2)

1044 (4:29) 3 "Eileen", Keith Richards **Main Offender** (Virgin 86499-2)

1045 (4:09) 1 "Never Aim To Please", Bash & Pop, **Friday Night Is Killing Me** (Sire 45133-2)

1131 (4:10) 4 "Don't Tear Me Up", Mick Jagger, **Wandering Spirit** (Atlantic 7 842436 2)

ALTERNATIVE ROCK- 25 24:48

1047 (3:14) 1 "Uh Huh Oh Yeh", Paul Weller, **Paul Weller** (Go! Discs 828 343-2)

1102 (3:18) 7 "If I Can't Change Your Mind", Sugar, **Copper Blue** (Rykodisc RCD 10239)

1048 (1:45) 11 "Lime House", The Breeders, **Pod** (4AD/Elektra 9 61331-2)

1043 (5:14) 10 "Man On The Moon", R.E.M., **Automatic For The People** (Warner Bros. 9 45055-2)

1063 (3:15) 4 "99.9 F", Suzanne Vega, **99.9 F** (A&M 31454-0005-2)

1061 (3:04) 9 "Candy Everyone Wants", 10,000 Maniacs, **Our Time in Eden**, (Elektra 961385 2)

1141 (4:00) 1 "Funky Ceili", Black 47, **Black 47** (SBK K2 7 80971-2)

NEW AGE- 26 26:39

1092 (4:00) 2 "Caribbean Blue", Enya, **Shepard's Moon** (Reprise 9 26775-2)

1127 (8:37) 4 "Lanzarote", Brian Eno, **The Shutov Assembly** (Opal 9 45010-2)

1098 (4:37) 1 "Lord Of The Wind", Kitaro, **Tunhuang** (Kuckuck CD 058)

1091 (3:47) 4 "So Long My Friend", Yanni, **Dare To Dream** (Private Music 01005-82096-2)

1145 (5:00) 12 "Roaring Of The Bliss", Tangerine Dream, **The Private Music of Tangerine Dream** (Private Music 01005-82105-2)

BROADWAY'S BEST- 27 29:22

1093 (4:19) 3 "If I loved You", Carousel Original Cast, **Carousel** (MCA MCAD 10048)

1093 (3:36) 5 "June is Busting Out All Over," Carousel Original Cast, **Carousel** (MCA MCAD 10048)

1093 (3:46) 7 "A Real Nice Clam Bake," Carousel Original Cast, **Carousel** (MCA MCAD 10048)

1155 (4:59) 11 "Master of The House," Les Miserables Original London Cast, **Les Miserables** (Encore CD1)

1155 (2:24) 15 "Little People," Les Miserables Original London Cast, **Les Miserables** (Encore CD1)

1155 (2:48) 12 "Wedding Chorale," Les Miserables Original London Cast, **Les Miserables** (Encore CD1)

1155 (2:16) 13 "Beggars at the Feast," Les Miserables Original London Cast, **Les Miserables** (Encore CD1)

1155 (4:39) 14 "Finale," Les Miserables Original London Cast, **Les Miserables** (Encore CD1)

<u>GOSPEL 28</u> 30:28

1116 (10:48) 7 "Amazing Grace", Aretha Franklin, **Amazing Grace** (Atlantic 906)

1121 (4:13) 2 "How I Got Over", Mahalia Jackson, **Jubilation! Great Gospel Performances, Vol. 2** (Rhino R2 70289)

1103 (4:58) 5 "Testify", Sounds Of Blackness, **The Evolution Of Gospel**, (Perspective 28968 1000-2)

1121 (3:56) 18 "The Love Of God", Rev. James Cleveland, **Jubilation! Great Gospel Performances, Vol. 2** (Rhino R2 70289)

1099 (3:34) 10 "None But The Righteous", Al Green, **One In A Million** (Word/Epic EK 77000)

1121 (2:59) 13 "Uncloudy Day", Staples Singers, **Jubilation! Great Gospel Performances, Vol. 2** (Rhino R2 70289)

<u>CHILDREN'S ENTERTAINMENT- 29</u>

1078 (3:05) 2 "This Old Man", Bob Dylan, **For Our Children** (Disney 60616-2)

1078 (3:38) 6 "Itsy Bitsy Spider", Little Richard, **For Our Children** (Disney 60616-2)

1128 (3:38) 8 "Three Little Birds", Freddie McGregor, **Reggae For Kids** (RAS 3095CD)

1078 (2:49) 11	"Getting To Know You", James Taylor, **For Our Children** (Disney 60616-2)
1128 (2:41) 1	"Puff The Magic Dragon", Gregory Isaacs, **Reggae For Kids** (RAS3095CD)
1078 (3:31) 4	"Mary Had A Little Lamb", Paul McCartney, **For Our Children** (Disney 60616-2)
1078 (1:42) 7	"Chicken Lips And Lizard Hips", Bruce Springsteen, **For Our Children** (Disney 60616-2)
1078 (2:42) 8	"Country Feelin's", Brian Wilson, **For Our Children** (Disney 60616-2)

<u>WORLD BEAT- 30</u> 28:08

1049 (4:52) 5	"Nicky B.B.", Matchatcha, **Matchatcha** (Africmusic CD 371)
1050 (4:08) 16	"Djobi Djoba", Gypsy Kings, **Live** (Elektra Musician 613902)
1051 (3:54) 3	"Ohude Manikiniki", Umahlathini Nabo, **Indestructable Beat Of Soweto** (Shanachie 43033)
1052 (5:22) 4	"Desert Rain", Outback, **Dance The Devil Away** (Hannibal HNCD 1369)
1053 (6:02) 1	"Moyo Wangu", Thomas Mapfumo, **Corruption** (Mango CCD9848)
1180 (3:30) 1	"After Hours," Battlefield Band, **After Hours**, (Temple COMD 2001)

ATTACHMENT A5. AT&T MUSIC COMPRESSION ALGORITHM

MUSIC PROGRAMMING

CD RADIO SYSTEM DEMONSTRATION

PREPARED BY

SPOTLIGHT MARKETING
3450 N. LAKE SHORE DRIVE SUITE 2009
CHICAGO, IL 60657

312 929 8230

Signal Compression: Technology Targets and Research Directions

Nikil Jayant, *Fellow, IEEE*

(Invited Paper)

Abstract—**Recent years have witnessed significant progress in the compression of speech, audio, image, and video signals. This has been the result of an interworking of coding theory, signal processing, and psychophysics. Advances in compression have been accompanied by numerous standards for digital coding and communication, both national and international. Emerging technology goals require even greater levels of signal compression. As we address these goals, and as we seek to define and approach fundamental limits in coding, we are guided by several promising trends in compression research, some of which represent cross-disciplinary evolution in the seemingly very different domains of acoustic and visual signals. Important aspects of new work in the field are the creation of refined methodologies for the measurement of signal distortion and coded signal quality, and increased interaction of source coding technology with other communications disciplines such as channel coding and networking.**

I. Introduction

SIGNAL coding is the process of representing an information signal in a way that realizes a desired communications objective such as analog-to-digital conversion, low bit-rate transmission, or message encryption. In the literature, the terms *source coding*, *digital coding*, *data compression*, *bandwidth compression*, and *signal compression* are all used to connote techniques used for achieving a compact digital representation of a signal, including the important subclasses of analog signals such as speech, audio, and image. The subject of this paper is the art and science of signal compression. When the terms *coding*, *encoding*, and *decoding* are used in this paper, they will all refer to the specific common objective of compression. An important theme of our discussion is the human receiver at the end of the communication process (Fig. 1).

Fig. 2 defines the role of signal compression (source coding) in digital communication. While the source coder attempts to minimize the necessary bit rate for faithfully representing the input signal, the *mo*dulator–*dem*odulator (modem) seeks to maximize the bit rate that can be supported in a given channel or storage medium without causing an unacceptable level p_e of bit error probability. The bit rate in source coding is measured in bits per sample or bits per second (b/s). In modulation, it is measured in bits per second per Hertz (b/s/Hz). The channel coding boxes add redundancy to the encoder bit stream for the purpose of error protection. In so-called coded modulation systems, the operations of channel coding and modulation are integrated for greater overall efficiency. The processes of source and channel coding can also be integrated in ways that will be illustrated in the last section of this paper.

The capability of signal compression has been central to the technologies of robust long-distance communication, high-quality signal storage, and message encryption. Compression continues to be a key technology in communications in spite of the promise of optical transmission media of relatively unlimited bandwidth. This is because of our continued and, in fact, increasing need to use bandlimited media such as radio and satellite links, and bit-rate-limited storage media such as CD-ROM's and miniaturized memory modules.

A. Background

The information-theoretical foundations of signal compression date back to the seminal work of Shannon [129], [130]. His mathematical exposition defined the information content or *entropy* of a source and showed that the source could be coded with zero error if the encoder used a transmission rate equal to or greater than the entropy and further, if it used a long processing delay, tending in general to infinity. In the special case of the infinite-alphabet or analog source, the encoding error tends to approach zero only at an infinite bit rate. However, in practice, the error is close enough to zero at finite rates. In the case of a finite-alphabet or discrete-amplitude source, the entropy is finite, and the bit rate needed for zero encoding error is finite as well. An important example of a finite-entropy source is an analog signal stored in a computer as a sequence of discrete amplitudes. The raw (uncompressed) bit rate of such a signal is typically 8, 16, or 24 bits per sample, respectively, for a grey-level image, high-quality speech (or audio), and a color image with three 8-bit components. The entropy, or the minimum bit rate for zero encoding error, will be smaller because of the statistical redundancy in the input sequence.

The inadequacies of the classical source coding theory are twofold. First, the theory is nonconstructive, offering

Manuscript received October 15, 1991; revised January 20, 1992.

The author is with the Signal Processing Research Department, AT&T Bell Laboratories, Murray Hill, NJ 07974.

IEEE Log Number 9107597.

0733-8716/92$03.00 © 1992 IEEE

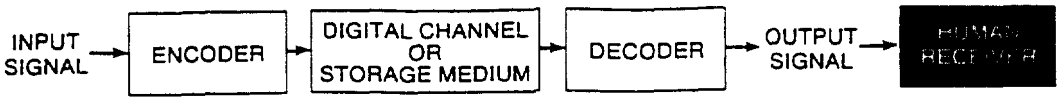

Fig. 1. Digital coding for signal compression.

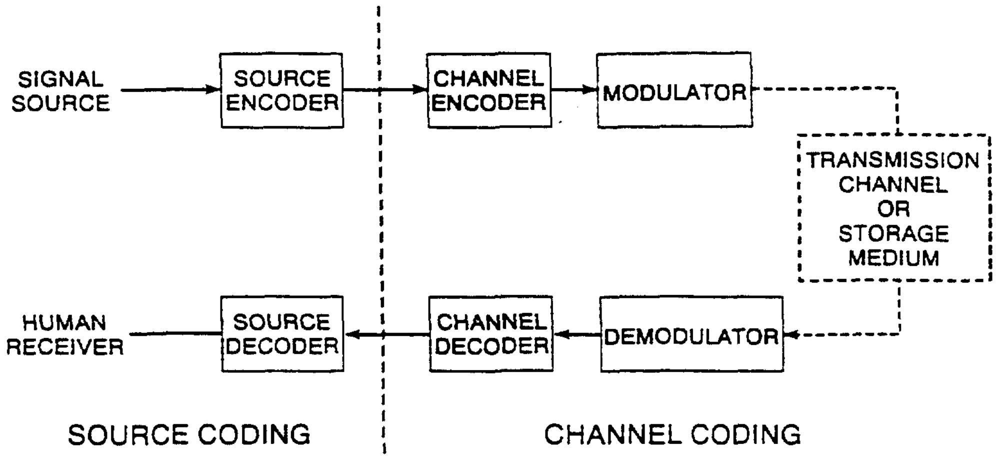

Fig. 2. Block diagram of a digital communication system.

bounds on distortion-rate performance rather than techniques for achieving these targets. We note, however, that the theory does anticipate important qualitative recipes such as the benefits of delayed encoding algorithms, as in vector quantization with a large vector dimension or block length. Second, the source model used in the classical theory does not capture what are now recognized as fundamental nuances in speech, audio and visual signals. These include the facts that the input signal is non-Gaussian, nonstationary, in general has a complex and intractable power spectrum, and the human receiver does not employ a mean-squared-error criterion in judging the similarity of a coded signal to the uncoded signal. As a result of the above complications, some of the observations of classical source coding do not carry over in an obvious way to signal compression as discussed in this paper. One such result is that the source entropy measured with a perceptual distortion criterion is different from, and generally much lower than, the classical entropy measured with a mean-squared-error criterion for coding distortion. Another classical result that needs to be reexamined is the thesis that, in principle, the processes of source and channel coding can be separated without loss of optimality. This result would hold very nicely for the digital communication of data sequences, but does not necessarily suggest an optimal solution to the complex problem of communicating audio or visual signals over a noisy channel with high perceptual fidelity, robustness, and bit-rate efficiency.

The technology and literature of signal compression have therefore evolved somewhat independently—with some valuable inspiration from Shannon's theory and the rate-distortion theory that followed it [5], [23], [36], [43], [54], [156], but really with a great deal of innovative engineering on the part of scientists closely familiar with the signals in question [1], [3], [4], [18], [19], [30], [32], [35], [42], [44], [47], [49], [53], [54], [59], [60], [62], [78], [80], [89], [90], [98], [100], [101], [102], [108], [113], [116], [121], [127], [132], [133], [135], [139], [151], [152], [154], [157]. In particular, work on speech compression has benefited greatly from studies of speech production and speech perception by humans, and research on visual perception has similarly impacted the parallel field of image compression. Although the mathematical theory of source coding is a common denominator, the fields of speech and image coding have been generally discussed by different schools, with a few recent exceptions. One of the purposes of this paper is to point out that as we address future technology targets in the disciplines of speech and image compression, common threads continue to exist. One such commonality is the increasing importance of matching the compression algorithm to the human perceptual mechanism—the auditory process in one case and the visual process in the other, leading to newly emerging generic techniques for quantization and time-frequency analysis in support of perceptually tuned coding, or *perceptual coding* for short.

B. The Dimensions of Performance in Signal Compression

The generic problem in signal compression is to minimize the bit rate in the digital representation of the signal while maintaining required levels of signal quality, complexity of implementation, and communication delay. We will now provide brief descriptions of the above parameters of performance.

Signal Quality: Perceived signal quality is often measured on a five-point scale that is well known as the *mean opinion score* or *mos* scale in speech quality testing: an average over a large number of speech inputs, speakers, and listeners evaluating the signal quality [22], [46], [62], [73], [74]. The five points of quality are associated with a set of standardized adjectival descriptions: *bad*, *poor*, *fair*, *good*, and *excellent*, and every example of an input being evaluated is assigned one of these levels in the course of a subjective test. Five-point scales of quality have also been used in image and audio testing, sometimes in the form of an inverted scale that categorizes levels of impairment [8] (*very annoying*, *annoying*, *slightly annoying*, *perceptible but not annoying*, and *imperceptible*) rather than quality. Other variations include the notion of averaging measurements over a selected *difficult* subset of input signals [136], in order to provide conservative scores of coder performance. In this paper, we use the original notion stated in this paragraph, a quality (rather than an impairment) scale, and an average over a large set of typical inputs, in each of the four generic categories: telephone-bandwidth speech, 20 kHz-bandwidth audio, still images, and video.

Our quantitative discussion of image and video quality will be impressionistic at best, given the multidimensionality of the problem (dependence of subjective quality on input scene, picture resolution, image size, and viewing distance) and the general lack of formal quality assessments in recent image-coding literature. In the field of speech coding, *mos* evaluations are well accepted and sometimes supplemented with measurements of *intelligibility* [146] and *acceptability* [145].

Bit Rate: We measure the bit rate of the digital representation in *bits per sample*, *bits per pixel* (b/p), or *bits per second* (b/s) depending on context, where *pixel* (sometimes shortened to *pel*) refers to a picture element or an image sample. The rate in bits per second is merely the product of the sampling rate and the number of bits per sample. The sampling rate is typically slightly higher than about twice the respective signal bandwidth, as required by the Nyquist sampling theorem [62].

Table I defines four commonly used grades of audio bandwidth. Typical sampling rates are 8 kHz for telephone speech, 16 kHz for AM-radio-grade audio, 32 kHz for FM-audio, and 44.1 or 48 kHz for CD (compact-disk) audio or DAT (digital audio tape) audio, both of which are signals of 20 kHz bandwidth. Respective bandwidths are strictly lower than half the corresponding sampling rates, following the principle of Nyquist sampling.

Table II defines commonly used grades of video in terms of sampling rate in pixels per second (p/s) or Hertz (Hz). The sampling rates for the CIF, CCIR, and HDTV formats defined in the table are 3, 12, and 60 MHz. Respective Nyquist bandwidths are approximately 1.5, 6, and 30 MHz, although bandwidth-limiting of image and video signals is in general less formal than the bandlimiting operations used for speech and audio signals. The HDTV format in the table is merely a specific example, one of several alternative formats. The sampling rates in the table refer to luminance information. Overheads for including color information are system-dependent.

TABLE I
DIGITAL AUDIO FORMATS

Format	Sampling Rate (kHz)	Bandwidth (kHz)	Frequency Band
Telephony	8	3.2	(200–3400 Hz)
Teleconferencing	16	7	(50–7000 Hz)
Compact Disk (CD)	44.1	20	(20–20000 Hz)
Digital Audio Tape (DAT)	48	20	(20–20000 Hz)

TABLE II
DIGITAL TELEVISION FORMATS

Format	Spatio-Temporal Resolution	Sampling Rate
CIF	360 × 288 × 30 =	3 MHz
CCIR	720 × 576 × 30 =	12 MHz
HDTV	1280 × 720 × 60 =	60 MHz

CIF: Common Intermediate Format
CCIR: International Consultative Committee for Radio
HDTV: One Example of a High Definition Television Format

In the CIF format, the color overhead is 50% in sampling rate, corresponding to a 50% subsampling relative to luminance in each of the horizontal and vertical directions, and a total of two chrominance components. Higher degrees of subsampling are sometimes used, leading to overall color overheads lower than 50%. In the line-interlaced CCIR format, the subsampling of color is performed only in the horizontal direction, and the final overhead in sampling rate is 100%.

Complexity: The complexity of a coding algorithm is the computational effort required to implement the encoding and decoding processes in signal processing hardware, and it is typically measured in terms of arithmetic capability and memory requirement. Coding algorithms of significant complexity are currently being implemented in real time, some of them on single-chip processors. Other related measures of coding complexity are the physical size of the encoder, decoder, or codec (*enc*oder plus *dec*oder), their cost (in dollars), and the power consumption (in watts or milliwatts, mW), a particularly important criterion for portable systems [15], [134].

Fig. 3 defines the evolution of digital signal processing (DSP) technology in terms of the number of instructions per second (mi/s) on a single general-purpose processor, as a function of time. The evolution is exponential, with no evidence of saturation in the near-term [86]. In the five-year period from 1990 to 1995, the typical per-chip capability is expected to increase tenfold, from about 25 mi/s to about 250 mi/s per chip. Supporting this evolution in arithmetic capability is a parallel advance in memory capability. The significance of the above advances is that sophisticated compression algorithms that demand increasing levels of complexity will be supported by DSP technology in the form of single-chip, multichannel im-

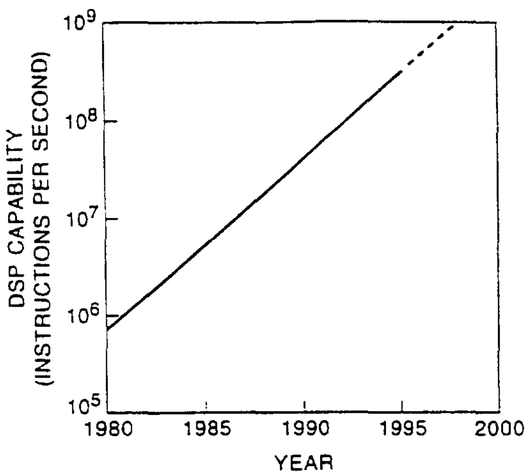

Fig. 3. Evolution of arithmetic capabilities in digital signal processing.

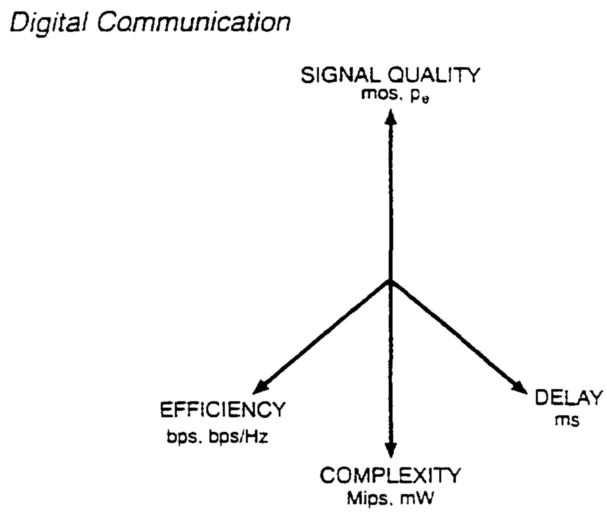

Fig. 4. The dimensions of coder performance.

plementations of the relatively less complex algorithms and realistic parallel-processing machines for the more complex techniques. Power dissipation and cost are also expected to decrease steadily, making DSP technology increasingly useful for personal portable devices and for other high-volume consumer applications.

Communication Delay: Increasing complexity in a coding algorithm is often associated with increased processing delays in the encoder and decoder. Although improved DSP capability can be used as an argument in favor of more sophisticated algorithms, the need to constrain communication delay should not be underemphasized. This need places important practical restrictions on the permissible sophistication of a signal compression algorithm. Depending on the communication environment, the permissible total delay for one-way communication (coding plus decoding delay) can be as low as about 1 ms (as in network telephony under conditions of no echo control) and as high as about 500 ms (as in very low bit-rate videotelephony (or *videophony*) where the delay performance is severely compromised in the interest of obtaining a received picture good enough for communication). Communication delay is largely irrelevant for applications involving one-way communication (as in television broadcasting) or storage and message-forwarding (as in voice mail).

C. Coding and Digital Communication

Fig. 4 describes performance criteria in digital communication by recapitulating the four dimensions of coder performance, explained specifically for source coding in Section I-B. These dimensions of performance apply to channel coding and modulation as well, although the units of quality and bit rate are different. Respective units, defined either in Section I-B or in the second paragraph of the paper, appear along the *quality* and *efficiency* axes in Fig. 4. Along each axis, the left and right entries refer to source and channel coding, respectively. The units of *complexity* and *delay* are identical for source and channel coding, although those parameters are used for different reasons in the two cases. Processing delay is used in source coding to remove signal redundancy. In channel coding, delay may be used for adding error protection bits and for processes such as interleaving for the randomization of burst errors.

The axes in Fig. 4 define a four-dimensional space in which some regions are theoretically allowable, and some regions are desirable for specific communication applications. Researchers in source and channel coding attempt to describe the allowable regions and tradeoffs as quantitatively as possible. The focus of this paper is on the domain of source coding. In particular, we shall comment extensively on the quantitative relationship between compressed signal quality and bit rate.

D. Road Map

In the remainder of this paper, we discuss applications and technology goals (Section II), the current tools of the trade (Section III), and emerging research directions (Section IV). The last section includes a brief discussion of efficient transmission, channel error protection, combined source and channel coding, and networking.

II. Applications, Standards, and Technology Goals

A. Applications

Fig. 5 depicts various applications of signal compression. The vertical axis in the figure does not have any special meaning. The numbers on the horizontal axis are bit rates *after* compression. The labels in the figure represent, in an approximate fashion, current capabilities. The bit rates spanned by these labels and, in some cases, the bit rates on which the labels are centered, represent the rates at which compressed signals render the corresponding application practical. As our capabilities in compression improve, the labels in the figure tend to drift

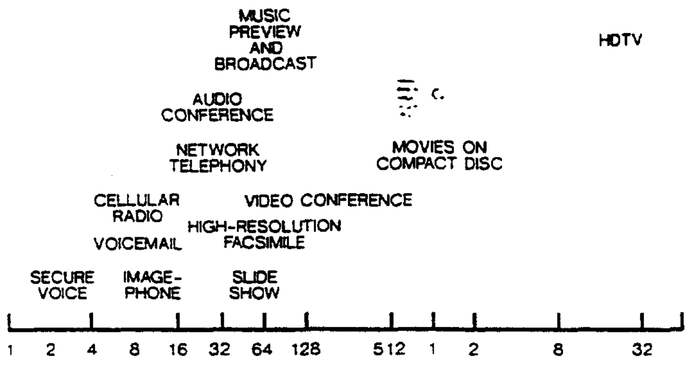

Fig. 5. Applications of signal compression.

to the left. Signals covered in Fig. 5 include telephone speech, wideband audio (speech and music), and a wide range of image signals, including still pictures and motion video. In the following paragraphs, we provide very brief summaries of current capabilities in the compression of speech (Fig. 6), audio (Fig. 7), and image (Fig. 8) signals.

Telephone Speech: Speech compressed to 2.4 kb/s provides a high level of intelligibility. However, speech quality, naturalness, and speaker recognizability are all poor at this bit rate. The need for digital encryption over a very wide range of transmission media is the main reason why 2.4 kb/s speech is widely used, particularly in government and defense communications [139].

A bit rate of 4.8 kb/s is sufficient to provide measurable gains in naturalness and speaker recognition. This bit rate is also of interest to government and defense applications [71]. With the increased demands of mobile telephony over bandlimited channels, 4.8 kb/s speech coding is also becoming very important for commercial communications using digital cellular radio. At 8 kb/s, which is the bit rate chosen for first-generation digital cellular telephony in North America [41], speech quality is high although significantly lower than that of the uncoded telephone-band signal.

At 16 kb/s and beyond, the speech quality is extremely close to that of the original, especially after a single stage of encoding and decoding. We use the term *network quality* to signify a performance level at which there is sufficient margin for additional functions such as multiple stages of encoding and decoding for speech, as well as high-accuracy transmission of nonspeech voiceband signals such as modem waveforms. The bit rate required for network-quality telephony, which was historically 64 kb/s and later reduced to 32 kb/s [9], is now coming down to 16 kb/s [13].

The application of *voicemail* involves speech storage as well as speech transmission for forwarding the voice message. Depending on the network environment used for this service and the desired speech quality in the received message, the bit rate can range from 4 to 32 kb/s, with increased focus expected in the 4 to 16 kb/s range in the future.

Wideband Speech: The 7 kHz signal has a higher voice quality than traditional telephony. This is partly due to increases in speaker presence and the naturalness of speech, as provided by the low-frequency enhancement (the added band from 200 to 50 Hz; see Table I), and partly due to increased intelligibility and crispness provided by high-frequency enhancement (the added band from 3400 to 7000 Hz). The higher quality of wideband speech is desirable for the extended communication task of a long audioconference call. It is also appropriate for other applications of loudspeaker telephony and for systems that include a high-quality speakerphone. It is also known that low-cost electret microphones can, in principle, support an incoming bandwidth of 7 kHz.

The standardized bit rate for high-quality coding of 7 kHz speech is 64 kb/s [10], [91], typically for an audioconference application using the Integrated Services Digital Network (ISDN). Recent algorithms have provided 7 kHz capability at 32 kb/s [115], permitting stereo-teleconferencing or dual-language programming over basic-rate ISDN. The projected capability for high-quality coding of 7 kHz speech is at least as low as 16 kb/s [34], [63], [64], [122].

The lower bit rates for wideband speech are also central to high-quality conferencing with combined audio and video. Current practice, at low values of total bit rate such as 64 kb/s, is to limit speech to the traditional telephone bandwidth of 3.2 kHz and to use 8 kb/s, or at the most, 16 kb/s for the coding of the audio channel. With advances in wideband speech compression, 16 kb/s coding of 7 kHz audio is expected to be an important component of conferencing, especially at higher ISDN rates such as 128 and 384 kb/s.

Wideband Audio: On a compact disk (CD), 20 kHz audio is sampled at 44.1 kHz and stored at 16 bits per sample or 706 kb/s per sound channel. Current algorithms for audio compression [6], [66], [97], [136], [137] provide CD-quality at 128 kb/s per channel for nearly all tested inputs, and CD-*like* quality at 64 kb/s per channel. These capabilities are important for emerging digital systems for audio broadcast and music preview. The capabilities are also central to applications that combine audio and visual functions, such as CD-ROM multimedia with a total bit rate of about 1.5 Mb/s and digital HDTV with a total bit rate of about 20 Mb/s (see Fig. 5).

Still Images: A 500 × 500 pixel color image, with the uncompressed format of 24 bits per pixel (b/p), will require about 100 seconds of transmission time over a 64 kb/s link. With 0.25 b/p coding, the transmission time is about 1 s, a number that would be deemed excellent for an interactive "slide show" [114]. Current technology for coding a 500 × 500 image is capable of providing good picture quality at 0.25 b/p for a wide class of color images, assuming a viewing distance of about six times the picture height [124], [147]. For most images, increasing the bit rate to 1 b/p provides excellent and, in some cases, perfect image quality. The corresponding transmission time over a 64 kb/s link is 4 s. Likewise, high-resolution facsimile typically takes several seconds of transmission

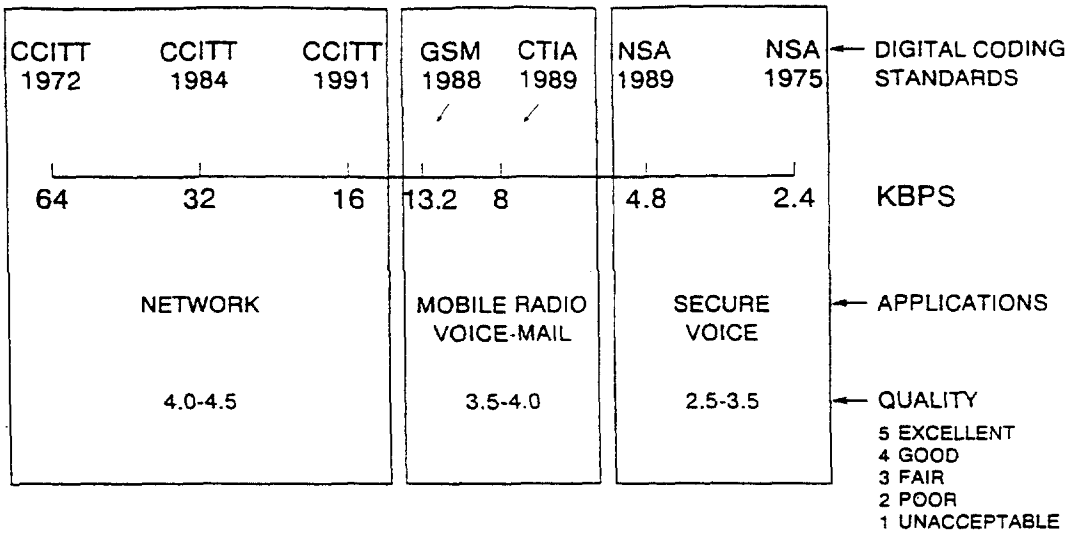

Fig. 6. Applications of telephone-speech compression, grades of digital speech quality, and standards for digital coding.

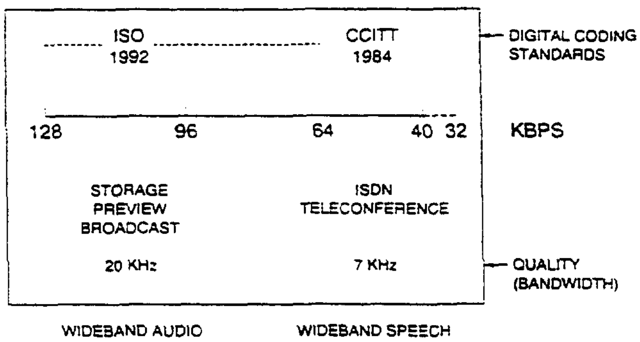

Fig. 7. Applications of wideband speech and audio, grades of bandwidth, and standards for digital coding.

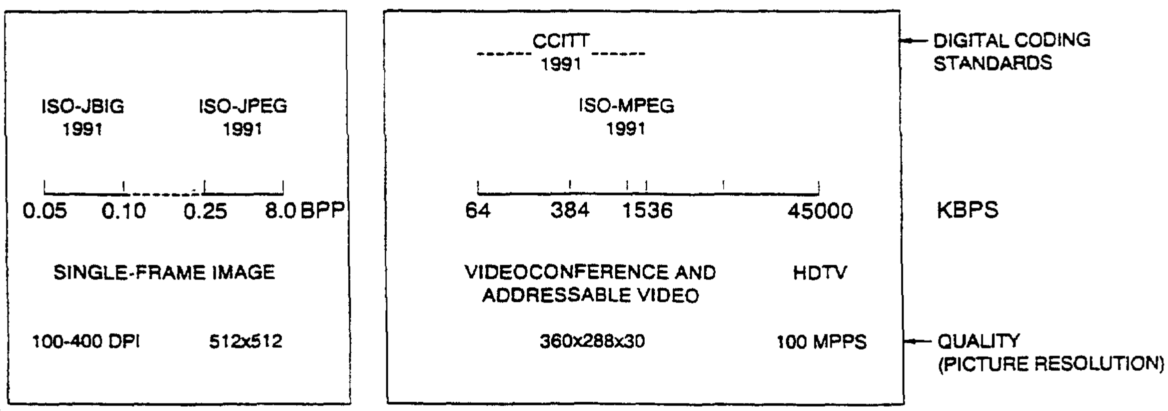

Fig. 8. Applications of image compression, grades of signal quality, and standards for digital coding.

time over a 64 kb/s channel even after the use of powerful techniques for fax compression [55], [56].

Techniques for *progressive transmission* [55], [147], [149] involves a first stage of coding characterized by a low bit rate and rapid picture access followed, if needed, by additional stages of transmission that upgrade the picture quality. Progressive transmission is ideal for applications such as telebrowsing. It is also appropriate for applications where one expects display modalities (terminals and printers) of varying resolutions.

The image-phone application in Fig. 5, which assumes the use of a telephone line and a 9.6 kb/s modem, involves pictures of very low spatial and temporal resolution. A typical resolution would be 100 × 100 pixels per frame, and about 3–6 frames per second. With an even lower temporal resolution, such as about 1 frame per second, the system degenerates to a sequence-of-stills service, sometimes referred to as *freeze-frame video*.

Digital Video: Serious videoconferencing requires CIF resolution (360 × 288 pixels per frame, Table II) or at least quarter-CIF resolution (180 × 144 pixels per frame). Input temporal resolutions are usually submultiples of 30 frames per second, say 15 or even 10, for bit rates lower than about 1.5 Mb/s. With CIF resolution and a bit rate of 1.5 Mb/s, the communications quality of the service is generally agreed to be high. With quarter-CIF or somewhat lower resolution and correspondingly lower values of temporal resolution, it is possible to achieve lower bit rates such as 48 or 112 kb/s. But the video quality is useful only if one accepts low levels of sharpness in the output picture and very low levels of motion activity in the input scene, as in the head-and-shoulders view of a single person—an environment sometimes referred to as video*telephony*, rather than video*conferencing*. The bit rates of 48 and 112 kb/s are appropriate for ISDN systems with total bit rates of 64 and 128 kb/s, respectively, and a bit rate of 16 kb/s for voice transmission. The bit rate of 384 kb/s is a very interesting number in the current state of technology. At this bit rate, it is possible to provide a fair, if not high, level of picture quality in the coding of a videoconference scene.

CD-ROM media have a net throughput rate of about 1.5 Mb/s for source data and a total bit capacity of a few gigabits. If video can be compressed to about 1 Mb/s, a CD-ROM device could store and play out about an hour or more of the video signal together with compressed stereo sound. This capability is central to various emerging applications of CD-ROM multimedia, including the specific example of a movie on an audio compact disk [81]. The additional capability of selecting a still-image snapshot of a desired part of the image sequence leads to the concept of *addressable video*. This is an important feature in emerging systems for video storage.

Uncompressed high-definition television (HDTV) has a bit rate of over a gigabit per second (the product of a sampling rate on the order of 60 MHz, as in Table II, and the representation of three color components with a total of 24 bits per sample). Compression of the HDTV signal to a bit rate on the order of a few tens of Mb/s will create several important opportunities for HDTV broadcasting. In particular, a bit rate in the range of 20 Mb/s will bring the service into the realm of a 6 MHz transmission channel [103], implying the capability of simulcasting the HDTV version of a program in vacant slots of an NTSC channel set. Transmission rates higher than 20 Mb/s are appropriate for higher-quality transmissions over satellite and broadband ISDN channels and applications of HDTV for movie production.

B. Compression Standards

The need to interoperate different realizations of signal encoding devices (transmitters) and signal decoding devices (receivers) has led to the formulation of several international and national standards for compression algorithms. Figs. 6–8 provide a nonexhaustive summary of compression standards for speech [9], [13], [41], [71], [78], [85], [139], audio [91], [97], [136], image [55], [56], [147], and video [81], [85], [103], [106]. Additional information provided in these illustrations includes typical applications, typical levels of signal quality, and the approximate date of formulating the standard.

The recent explosion in standards activity has had an important impact on research and development in the field. Standards have led to an increased focus in applied research. They have sometimes stimulated highly productive new research as well. They have also elevated the threshold of performance that a novel research algorithm needs to exceed before it is widely accepted, given that the supplanting of a recently endorsed standard is generally difficult and expensive.

Several applications of signal compression are decoder-intensive in the sense that users need access only to a decoder, the encoding being a one-time operation by the provider of the service. Examples are multimedia and HDTV decoders. In recognition of this, corresponding standards have specified the decoder algorithms and bit stream syntax rather than the encoder. In these cases, compatible enhancements to the standard are possible in the encoding module, as well as in optional modules of pre- and postprocessing—prefiltering at the encoder and postfiltering at the decoder.

C. Quality of the Compressed Signal

We have noted earlier that there are several dimensions defining the performance of a coding system. If we ignore the dimensions of algorithmic complexity and communication delay for the moment, coder improvements can be demonstrated in two ways: by measuring signal quality improvement at a specified bit rate, or by realizing a specified level of signal quality at a lower bit rate. Depending on the application, one of the above approaches would be more relevant than the other. For example, in the problems of coding telephone speech at 4 kb/s and HDTV at 15 to 20 Mb/s, the bit rates are defined by important generic applications and the goal of coding research is to enhance signal quality at those rates. On the other hand, in the field of digital audio broadcasting (DAB) where the signal quality needs to be transparent to the coding algorithm and equivalent, say, to that of CD-audio, the goal is to demonstrate such performance at progressively lower rates (say, at 48 or 64 kb/s per channel rather than at 96 or 128 kb/s per channel as in currently proven algorithms).

So far, our discussions of bit rate have been in terms of kilobits per second (kb/s) for speech and audio, and both kilobits per second (kb/s) and megabits per second (Mb/s) in the case of video. All of these numbers can obviously be converted to equivalent numbers in bits per sample based on sampling rates such as the illustrative numbers in Tables I and II. For example, the 4 kb/s, 48 kb/s, and 15 Mb/s rates for 8 kHz-sampled speech, 48 kHz-sampled audio, and 60 MHz-sampled HDTV correspond to 0.5, 1, and 0.25 bits per sample, respectively. In our ensuing definition of technology targets in signal compression, we shall use the normalized unit of bits per sample in the interest of a unified perspective for audio and visual signals.

D. Technology Targets

Fig. 9 is a simplified and impressionistic summary of current capabilities in signal coding, expressed in terms of subjective quality as a function of bit rate. The results are derived from a combination of published work [9], [10], [22], [91], unpublished reports [136], and by collective impressions of experts, especially in the case of image and video signals where formal evaluations of quality are not generally available. Signal quality is measured on a subjective five-point scale ranging from *bad* to *excellent*, as in our earlier description of the *mos* scale.

One of the implications in Fig. 9 is that video signals are the easiest to compress on a bit-per-sample basis. This is attributable to the well-known redundancy in video information in both the spatial and temporal domains of the signal, a property that is also reflected in the extreme lowpass nature of the power spectral density of typical video. By contrast, it is not unusual to encounter relatively flat power spectra in 20 kHz audio. This, combined with the universal expectation of very high levels of quality in entertainment audio, leads to generically lower subjective scores in compressed audio at a given number of bits per sample.

In the category of still images, facsimile documents constitute a special subclass if we agree to regard text and line graphics, rather than grey-level photographs, as typical fax documents. A half-toned (black–white) document is generally highly compressible. The bit rate for the lossless coding of a fax document can be typically on the order of 0.1 bit per sample.

As we seek to advance the state of the art as depicted earlier in Fig. 5, it is useful to talk about bit rate targets at which one expects the four signals in Fig. 9 to be digitized with a quality rating such as 4 or higher. Without loss of generality or realism, all of these targets can be collectively described by the bit-rate-independet horizontal broken line of 4.5-quality in Fig. 9. Clearly, we are closer to this goal in some signal domains than others. It is also possible that the 4.5-quality goal at rates down to 0.25 bit per sample is impossible to achieve in some cases, regardless of coder complexity or processing delay, because of fundamental limits imposed by information theory and the acuity of the human perceptual system. But it is fair to ask the question: as we seek to approach these (sometimes unattainable) levels of high quality at low bit rates, what are the techniques most likely to succeed?

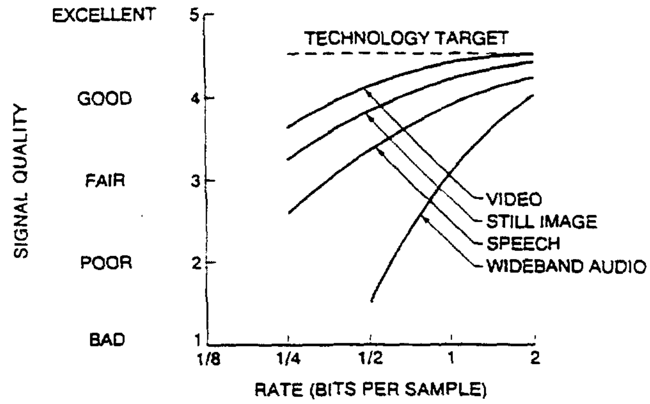

Fig. 9. Current capabilities in the coding of audiovisual information.

III. Tools of the Trade

There are three fundamental operations that are common to low bit-rate signal coding: reduction of signal redundancy in the input signal, removal of irrelevant information in the operation of quantization, and signal enhancement by postfiltering. Of these, postprocessing is generally considered to be a process outside of the coding operations *per se*, although the benefits of performing the process can be very significant, as in low bit rate speech coding [12], [117]. Prefiltering can, likewise, increase the performance of a compression algorithm. This is accomplished in video coding, for example, by the reduction of camera noise in the input image or by the insertion of an explicit bandlimiting filter. The remainder of this paper will focus on the two operations that are intrinsic to signal coding: *removal of redundancy* and *reduction of irrelevancy*.

Almost all sampled signals in coding are redundant because Nyquist sampling typically tends to preserve some degree of intersample correlation. This is reflected in the form of a nonflat power spectrum. Greater degrees of nonflatness, as resulting from a lowpass function for signal energy versus frequency or from resonances (in audio) and periodicities (in audio and video), lead to greater gains from redundancy removal. These gains are also referred to as prediction gains or transform coding gains, depending on whether the redundancy is processed in the time or frequency (or transform) domain.

In a signal compression algorithm, the inputs to the quantizing system are typically sequences of prediction error or transform coefficients. The idea is to quantize the time components of the prediction error or the transform coefficients just finely enough to render the resulting distortion imperceptible, although not mathematically zero. If the available bit rate is not sufficient to realize this kind of perceptual transparency, the intent is to minimize the perceptibility of the distortion by shaping it advantageously in time or frequency, so that as many of its components as possible are masked by the input signal itself. We use the term *perceptual coding* to signify the matching of the quantizer to the human auditory or visual system, with the goal of either minimizing perceived distortion or driving it to zero where possible.

The parts of a coder that process redundancy and irrelevancy are sometimes separate, as in the above explanation. On the other hand, there are examples where the two functions cannot be easily separated. One example is a vector quantizer that combines intersample processing and quantization in a single stage of processing.

Almost all coding systems depend on the complementary interworking of the two basic operations defined above. A notable exception is a pulse code modulation (PCM) system, based on memoryless coding and quantizing algorithms, where there is no attempt to remove signal redundancy. This simple procedure is adequate for high-quality coding at bit rates in the range of 8-16 bits per sample, depending on the input signal. On the other hand, low bit rate coders, such as those evaluated in Fig. 9, depend heavily on more sophisticated signal analysis, processing delay, and redundancy removal prior to perceptually tuned quantization.

Speech signals have a universal production model which provides a very powerful framework for redundancy removal. By contrast, audio and image signals, although often very structured and redundant, lack a universal model of signal production. In this sense, the role of perceptually-efficient quantization becomes even greater in the coding of such signals, especially at lower bit rates.

We now discuss some generic examples of coding algorithms, drawing from the fields of speech, audio, and image signals. Our attempt here is not to provide an exhaustive overview, but to provide a suitable background against which to portray some evolutionary trends in current coding literature. This discussion, in turn, will set the stage for the research directions described in Section IV.

A. Closed-Loop LPC Coding of Speech

The universal model of speech production, consisting of an excitation followed by a linear filter [Fig. 10(a)], has made possible the ubiquitous use of the linear predictive coding (LPC) model in various applications of low bit rate speech, prominently in the form of the LPC vocoder for the extremely low bit rate of 2.4 kb/s [139] and code-excited linear prediction (CELP) and related algorithms for bit rates such as 4.8, 6.4, 8, and 16 kb/s [3], [13], [71], [79], [131]. The lower bit rates of CELP are typical of cellular radio and satellite applications. The Regular Pulse Excitation algorithm (RPE) [78] used in the European digital cellular radio system at the bit rate of 13.2 kb/s, as well as algorithms such as multipulse speech coding [4], vector-adaptive predictive coding [12], and vector-sum-excited linear predictive coding (VSELP), are similar to CELP in that they are *closed-loop* LPC systems unlike the *open-loop* algorithm used in the LPC vocoder. The 8 kb/s VSELP system [41] is the basis of the first-generation North American digital cellular telephone standard. The 16 kb/s CELP coder, with a backward-

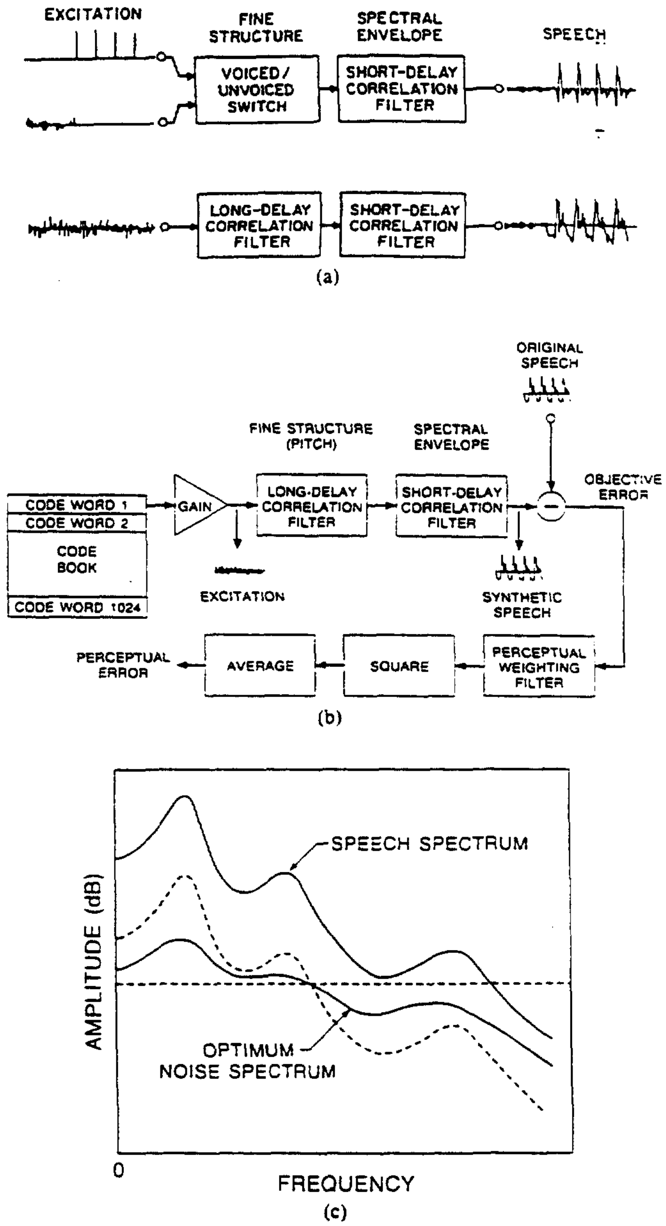

Fig. 10. (a) The excitation-filter model of speech production. (b) A block diagram of a code-book excited linear predictive coder (CELP). (c) Generic form of the optimally shaped spectrum of coding distortion (after [4]).

adaptive LPC predictor, is the basis for a network-quality speech coding system [13].

In the closed-loop systems, the basic model for speech synthesis is the traditional LPC scenario of an excitation signal driving an all-pole filter. However, the nature of the excitation is not decided by a simple binary categorization of speech into *voiced* and *unvoiced* segments (as in vocoding), but by an exhaustive search procedure that is reminiscent of waveform coding. This procedure picks, for each all-pole filter model, the best possible excitation signal for each speech segment (with a typical duration in the range of 5-20 ms) based on a frequency-weighted mean square error criterion. The closing of the *analysis-by-synthesis* loop provides the mechanism whereby a frequency-weighted mean squared error criterion is used to select the optimum excitation signal for the LPC synthesizer [Fig. 10(b)].

The closed-loop LPC method for speech coding is a very mature and successful paradigm. The process of redundancy removal is effected by the time-varying LPC filter which, at the encoder, acts as an LPC predictor. Perceptually-efficient quantization is provided by the weighted-error-steered selection of the excitation codevector (which is equivalent to vector quantization of the LPC residual). The voiced–unvoiced binary excitation codebook in LPC vocoding can be regarded as an extremely low bit rate version of that vector quantizer.

CELP and CELP-like methods are also very appropriate for the low bit rate coding of wideband speech. In currently reported work, with 7 kHz inputs, very high quality is obtained at 32 kb/s while very good communications quality is possible at 16 kb/s [63].

Fig. 10(c) illustrates the qualitative form of noise-shaping used in closed-loop LPC coding of speech. Coding distortion is least audible when its power spectrum is neither white nor speech-like (as shown by the broken characteristics in the figure), but has an intermediate form (the thinner solid characteristic in the figure). This prescription is a somewhat simplified translation of what we know about the way a stronger signal tends to mask a weaker signal in its frequency vicinity in the human auditory system. Complete masking of the weaker signal (coding distortion, in the current discussion) occurs when the signal-to-distortion ratio at each frequency equals or exceeds a critical threshold. In the current state of the art, low bit rate speech coders do not typically realize this ideal situation. Instead, they attempt to approach it using the simplified prescription of the intermediate noise spectrum depicted in Fig. 10(c). This kind of noise shaping is achieved in CELP and CELP-like systems by the process of selecting the excitation codevector that minimizes a frequency-weighted error for a given speech segment and the corresponding LPC filter.

B. *Perceptual Transform Coding of Wideband Audio*

Fig. 11(a) illustrates a key concept of perceptual coding that we shall call *just noticeable distortion* (JND, which stands for just noticeable *difference* in psychophysical literature). If the coding distortion is at or below the level of the staircase JND function at all frequencies, the input signal masks the distortion completely and transparent coding results, meaning that the compressed audio is indistinguishable from the original uncompressed audio. Important consequences of the above result are that: a) most signal components can be quantized fairly coarsely; and in particular, b) signal components below the JND can be completely discarded without causing any perceivable distortion. This is because the objective distortion at the discarded frequencies is equal to the (unsent) input magnitude and, if this is below the corresponding JND, the resulting degradation is perceptually irrelevant. A good example of discardable frequencies in Fig. 11(a) is the 5-to-6 kHz band. The combined result of discarding

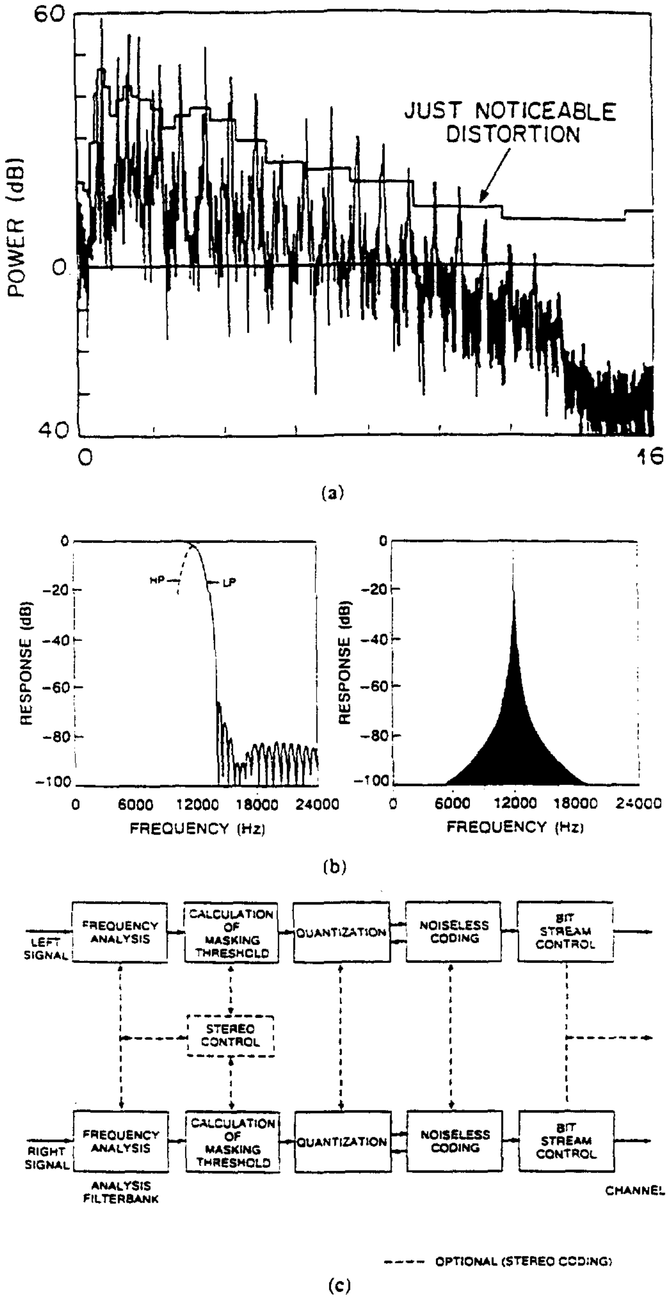

Fig. 11. (a) The just-noticeable distortion as a function of frequency. (b) Frequency response of QMF and MDCT filterbanks. (c) Block diagram of a perceptual coder capable of transparent coding (after [66]).

several frequencies and quantizing many others quite coarsely is that the average bit rate is very low, on the order of 2–3 bits per sample. These rates provide perceptually transparent coding of the 16 bit-per-sample input because of the use of the JND principle.

The JND profile is calculated from a short-term frequency description of the signal to be coded, using principles of human audition—in particular, detectability criteria and properties of the cochlea, the frequency-separation mechanism of the human ear. Quantitative descriptions of JND data are well understood for simple and idealized descriptions of masking and masked signals (for example, tones and noise). But in the general case of an